

Automatic Text Summarization using a Machine Learning Approach

Joel Larocca Neto, Alex A. Freitas, Celso A. A. Kaestner



Agenda

- Einführung in die automatische Textzusammenfassung
- Merkmale und Methoden
- Maschinelles Lernen in der Textzusammenfassung
- Experimentelle Ergebnisse und Diskussion
- Schlussfolgerungen und zukünftige Forschung

Einführung in die automatische Textzusammenfassung

- Reduzierung der Textgröße bei Erhaltung des Informationsgehalts
- Unterschied zwischen extraktiven und abstrakten Zusammenfassungen
- Vorteile der automatischen Zusammenfassung

Merkmale der Textzusammenfassung

- Statistische Merkmale wie Termfrequenz
- Linguistische Merkmale basierend auf Argumentstruktur
- Bedeutung der Satzposition und -länge

Methoden der Textzusammenfassung

- Vektorraum-Modell zur Ähnlichkeitsmessung
- Verwendung von Stop-Wort-Eliminierung und Stemming
- Klassifizierung von Sätzen basierend auf Merkmalen

Maschinelles Lernen in der Textzusammenfassung

- Einsatz von Naive Bayes und C4.5 als Klassifikatoren
- Training mit dokumentierten Referenzzusammenfassungen
- Klassifizierung als zweiklassiges Problem

Experimentelle Ergebnisse

- Vergleich von automatisch und manuell erstellten Zusammenfassungen
- Naive Bayes übertrifft C4.5 in der Leistung
- Verwendung von TIPSTER-Datenbank für Experimente

Diskussion der Ergebnisse

- Höhere Präzision und Recall bei höherer Kompressionsrate
- Bedeutung der Wahl des Klassifikators
- Vergleich mit nicht-trainierbaren Methoden

Schlussfolgerungen

- ML-Ansatz ermöglicht objektive Bewertung
- Naive Bayes als effektiver Klassifikator
- Notwendigkeit neuer Klassifikationsalgorithmen

Zukünftige Forschung

- Entwicklung spezialisierter Klassifikatoren
- Erweiterung der Merkmalsauswahl
- Verbesserung der linguistischen Verarbeitung

Herausforderungen

- Subjektive Bewertung von Zusammenfassungen
- Komplexität der natürlichen Sprachverarbeitung
- Dissonanz zwischen menschlichen und automatischen Zusammenfassungen

Zusammenfassung

- ML-basierte Ansätze bieten vielversprechende Ergebnisse
- Wichtigkeit der Merkmalsauswahl und Klassifikatorwahl
- Potenzial für weitere Verbesserungen in der Textzusammenfassung